

SOZLab

Linux GUI + CLI for solvent occupancy zone analysis

Version: 0.1.1 | Date: January 24, 2026

Table of Contents

SOZLab Tutorial	2
What SOZLab does	2
Quick start	3
CLI reference (current flags)	4
Output directory behavior	5
UI overview (layout and navigation)	5
Project step (inputs and run configuration)	6
Define step (SOZ builder tabs)	7
QC step	9
Explore step (interactive analysis)	9
Export step (data, report, extraction)	11
Plot export and PowerPoint guidance	12
Project JSON schema (current)	13
Outputs and file map	16
Extraction (GUI + CLI)	16
Reproducibility and QC	17
Troubleshooting (Symptom -> Confirm -> Fix)	17
Appendix: keyboard shortcuts	18

SOZLab Tutorial

What SOZLab does

SOZLab is a Linux GUI + CLI for solvent occupancy zone (SOZ) analysis of MD trajectories. It evaluates, per frame, which solvent residues satisfy a user-defined logic tree around one or more seed selections and reports occupancy, entry/exit events, residence statistics, plots, and extractable frame sets.

Operational definition of a SOZ

A SOZ is the set of solvent residues that satisfy a logic tree in a given frame. The evaluated set is residue-based (resindices/resids), not atoms. All occupancy counts and events refer to solvent residues, which avoids atom-count bias for multi-atom solvents.

Per-frame metrics (definitions)

- 1 **frame**: zero-based trajectory frame index after applying **analysis.frame_start**, **analysis.frame_stop**, and **analysis.stride**.
- 1 **time**: **ts.time** from MDAnalysis for the frame. Values are written to CSV as-is.
- 1 **n_solvent**: integer count of solvent residues in the SOZ for that frame.
- 1 **entries**: number of solvent residues that appear relative to the previous sampled frame.
- 1 **exits**: number of solvent residues that disappear relative to the previous sampled frame.

Stride effects (critical)

Entries and exits are computed between sampled frames. If **analysis.stride** is large, short-lived events can be missed and entry/exit counts only reflect changes between sampled frames.

Units and time bases

- 1 Length cutoffs accept **unit: "A"** or **unit: "nm"**. Internally, SOZLab converts to angstroms (1 nm = 10 Å).
- 1 Time comes directly from MDAnalysis. The GUI divides **time** by 1000 and labels it in ns (assumes ps input). If your trajectory time units are not ps, confirm the unit before interpreting plots.
- 1 GROMACS conventions use nm for length and ps for time. Always validate your expectations when switching between toolchains.

Quick start

GUI quick start (sample project)

1) Launch: **sozlab-gui**. 2) Click **Load** (header bar) and open **examples/sample_project.json**. 3) In the left drawer, review **Inputs**, **Analysis Window**, and **Output Settings**. 4) Click **Run**. The status bar shows progress and a log path. 5) Use the stepper tabs: **Project**, **Define**, **QC**, **Explore**, **Export**. 6) In **Explore**, inspect the Timeline and plots. Use **Export Data** or **Export Report** when ready.

CLI quick start (sample project)

```
sozlab-cli run --project examples/sample_project.json --output out --progress --report
```

Validation (fast sanity check):

```
sozlab-cli validate --project examples/sample_project.json --max-frames 200
```

Extraction example:

```
sozlab-cli extract --project examples/sample_project.json --soz shell_only --rule
"n_solvent>=1" --min-run 5 --gap 1 --out out/extracted
```

Example projects

- 1 **examples/sample_project.json**: minimal example.
- 1 **examples/gromacs_template_project.json**: GROMACS-oriented template.
- 1 **examples/soz_configs/**: individual SOZ definitions for Advanced JSON.

CLI reference (current flags)

Run

```
sozlab-cli run --project PROJECT [--output OUTPUT] [--stride STRIDE] [--start START]
[--stop STOP] [--no-ids] [--no-per-frame] [--progress] [--report]
```

- 1 **--project**: project JSON file (required).
- 1 **--output**: output directory for results and logs.
- 1 **--stride**: frame stride override.
- 1 **--start**: start frame index.
- 1 **--stop**: stop frame index (exclusive).
- 1 **--no-ids**: disable per-frame solvent IDs.
- 1 **--no-per-frame**: disable per-frame table export.
- 1 **--progress**: show progress bar in terminal.
- 1 **--report**: generate report after analysis.

Validate

```
sozlab-cli validate --project PROJECT [--max-frames MAX_FRAMES]
```

- 1 **--project**: project JSON file (required).
- 1 **--max-frames**: maximum frames to validate (default 200).

Extract

```
sozlab-cli extract --project PROJECT [--soz SOZ] [--rule RULE] [--min-run MIN_RUN] [--gap
GAP] [--out OUT] [--format FORMAT] [--prefix PREFIX] [--output OUTPUT]
```

- 1 **--project**: project JSON file (required).
- 1 **--soz**: SOZ name to filter.
- 1 **--rule**: rule string (e.g., **n_solvent>=1**).
- 1 **--min-run**: minimum consecutive frames.
- 1 **--gap**: gap tolerance to merge runs.

- 1 **--out**: extraction output directory.
- 1 **--format**: trajectory format (default **xtc**).
- 1 **--prefix**: output filename prefix (default **extracted**).
- 1 **--output**: override output dir for logs only.

Output directory behavior

- 1 **outputs.output_dir** is the base directory for logs and exported results.
- 1 GUI **Output Settings** edits **outputs.output_dir**. Export Data/Report prompts for a directory and updates this value.
- 1 CLI **run --output OUT** overrides **outputs.output_dir** for that run.
- 1 CLI **extract --output OUT** overrides only the log directory; **extract --out OUT_EXTRACT** controls extraction outputs.

UI overview (layout and navigation)

Header bar

- 1 Stepper: **Project**, **Define**, **QC**, **Explore**, **Export** (switches main pages).
- 1 Actions: **Load**, **Save**, **New SOZ**, **Quick**, **Run**, **Cancel**, **Export Data**, **Export Report**.
- 1 Toggles: **Console** (diagnostics console), **Drawer** (left panel), **Inspector** (right panel).
- 1 UI controls: **Theme** (Light/Dark), **Scale**, **Presentation** (larger UI, hides drawers).
- 1 Shortcuts: see the Appendix.
- 1 Command palette (**Ctrl+K**) opens a searchable list of common actions.

Action behavior:

- 1 **Load** opens a project JSON/YAML file.
- 1 **Save** writes the current project to JSON/YAML (prompts on first save).
- 1 **New SOZ** applies the current Wizard definition to the project.
- 1 **Quick** sets a coarse stride based on **analysis.preview_frames** and runs a fast preview.
- 1 **Run** enforces Project Doctor checks before analysis.
- 1 **Cancel** stops an in-progress run.

Main layout

- 1 Left drawer: project settings, output settings, Project Doctor, defined SOZ list.
- 1 Center page: content for the selected stepper page.
- 1 Right inspector (optional): context-sensitive status for the current page.

- 1 Bottom console (optional): recent log lines and quick refresh.
- 1 Status bar: run status messages and progress bar.

Inspector contents (when enabled):

- 1 Project: Provenance Stamp with copy button.
- 1 Define: live selection status tips.
- 1 QC: QC status summary.
- 1 Explore: active SOZ, metric, and time window.
- 1 Export: extraction status hints.

Project step (inputs and run configuration)

If no project is loaded, **Run** or **Quick** will prompt for topology and optional trajectory, then create a default project in memory.

Inputs (left drawer)

- 1 **Topology**: label shows current file; **Change** opens a file dialog.
- 1 **Trajectory**: label shows current file; **Change** and **Clear** manage it.
- 1 **Metadata**: summary line (SOZ count, selection count, stride, frames if available).

Analysis Window (left drawer)

- 1 **Frame start**: integer start frame index.
- 1 **Frame stop**: integer stop frame index; use **End** to include full trajectory.
- 1 **Stride**: sampling stride (integer ≥ 1).

Output Settings (left drawer)

- 1 **Output dir**: base directory for analysis outputs and logs.
- 1 **Report format**: **html** or **md**.
- 1 **Write per-frame CSV**: controls whether **per_frame.csv** is written.
- 1 **Write parquet**: writes **.parquet** alongside CSV (requires **pyarrow**).

Project Doctor (left drawer)

- 1 **Run Project Doctor**: preflight checks on inputs, selections, solvent, probe, units, and PBC.
- 1 **PBC Helper**: shows recommended **gmx trjconv** commands for preprocessing.
- 1 Selection table: lists selection labels, counts, uniqueness requirements, and suggestions (columns: Selection label, Count, Require unique match, Expect count, Selection, Suggestions).

Defined SOZs (left drawer)

- 1 List of SOZ definitions currently in the project file.

Project page (center)

- 1 **Project Summary:** topology/trajectory details and PBC status.
- 1 **Run Summary:** summarized results after analysis; **Show raw JSON** reveals raw summary.

Define step (SOZ builder tabs)

Wizard tab (main builder)

Fields and controls:

- 1 **SOZ name:** name for the SOZ definition.
- 1 **Solvent label:** display label only (does not change selection logic).
- 1 **Solvent resnames:** comma-separated residue names used to identify solvent molecules.
- 1 **Probe selection:** MDAnalysis selection applied within solvent residues.
- 1 **Probe position:** **atom**, **com**, or **cog**.
- 1 **Include ions + Ion resnames:** optional ion residues included as solvent.
- 1 **Selection A + Selection A unique match.**
- 1 **Shell cutoffs (A):** comma-separated values in angstroms.
- 1 **SOZ probe mode:** **probe**, **atom**, **com**, **cog**, or **all**.
- 1 **Selection B (optional) + Selection B unique match.**
- 1 **Selection B cutoff (A).**
- 1 **Selection B combine:** **AND** or **OR**.
- 1 **Explain my SOZ:** a readable summary of the current wizard definition.
- 1 **Preview:** evaluates the SOZ for frame 0 (requires inputs).

Notes:

- 1 The Wizard uses angstrom inputs only. To use **unit: "nm"**, edit the SOZ JSON in **Advanced**.
- 1 **Probe position = com** requires atom masses in the topology; otherwise, use **cog**.
- 1 The Wizard stores the probe in **solvent.probe** and keeps legacy **water_oxygen_names** in sync.

Selection match preview (Wizard)

- 1 **Live:** update when inputs change.
- 1 **Use trajectory:** include trajectory frames, not just topology.
- 1 **Max rows:** limit output rows.

- 1 **Preview:** Selection A or Selection B.
- 1 Status lines show match counts and suggestions (columns: Atom index, Atom name, Residue name, Residue number, Segment ID, Chain ID, Molecule type).

Selection Builder tab

Builds MDAnalysis selection strings:

- 1 **Scope:** all atoms, protein, backbone, sidechain, nucleic.
- 1 **Resname, Resid/resnum, Atom name, Segid, ChainID.**
- 1 **Build** generates the selection string.
- 1 **Use as Selection A/B** sends to Wizard.
- 1 **Send to Tester** pushes to Selection Tester.

Bridges tab

Defines solvent bridges between two selections.

- 1 Table columns: **name, selection_a, selection_b, cutoff_a, cutoff_b, unit, probe_mode.**
- 1 **Add Bridge, Remove Selected.**
- 1 Each bridge produces **bridge_<name>/** outputs.

Residue Hydration tab

Counts residue contacts with solvent in the SOZ.

- 1 Table columns: **name, residue_selection, cutoff, unit, probe_mode, soz_name.**
- 1 **hydration_freq** is the fraction of frames where the residue contacts the solvent set.
- 1 If **soz_name** is empty, the first SOZ is used.

Advanced tab

- 1 Paste a SOZ JSON definition and click **Apply JSON to SOZ list.**
- 1 Use this for multi-branch trees, non-A units, or advanced probe modes.

Selection Tester tab

- 1 **Test Selection:** evaluates a selection string.
- 1 **Test Probe:** evaluates the current solvent probe selection.
- 1 **Max rows** and **Load trajectory** control sampling.
- 1 Results list atom/residue identifiers and suggestions (columns: index, resname, resid, resnum, segid, chainID, moltype, name).

QC step

QC Summary page

- 1 **QC Summary**: a readable summary of preflight and run-time warnings.
- 1 **Show raw QC JSON**: shows raw QC payload.
- 1 **Diagnostics** tab: log viewer (same controls as the Logs tab).

Diagnostics and logs

- 1 **Log path, Refresh, Open Log, Copy Errors**.
- 1 Filters: Level (All/INFO/WARNING/ERROR), search box, **Collapse** **tracebacks**.

Explore step (interactive analysis)

Timeline top area (tabbed)

The top plot region is tabbed: **Timeline** and **Entry/Exit** share the same plot area.

Timeline tab (Occupancy Timeline)

Controls:

- 1 **SOZ** selector.
- 1 **Metric** and **Secondary** (secondary plots on right axis).
- 1 **Smooth + Window** (rolling mean in frames).
- 1 Display toggles: **Overlay all SOZs**, **Step plot**, **Markers**, **Clamp $y \geq 0$** , **Mean line**, **Median line**, **Shade occupancy** (threshold), **Brush time window**.
- 1 **Clear brush** resets the time window.
- 1 Actions: **Save Plot**, **Copy** (clipboard), **Export** (file dialog).
- 1 **Compute Timeline Statistics** calculates summary metrics (mean, median, entries/exits, residence stats).

Axes:

- 1 X-axis: time in ns (GUI divides **time** by 1000).
- 1 Y-axis: selected metric (count or fraction depending on metric).

Metric meanings:

- 1 **n_solvent**, **entries**, **exits** are counts per frame.
- 1 **occupancy_fraction** is 0/1 per frame (1 when **n_solvent**>0).

Brush behavior:

- 1 When enabled, the brush time window filters Histogram, Heatmap, Event Raster, and tables.

Entry/Exit tab (Entry/Exit Rate)

Controls:

- 1 **Entry/Exit mode**: Events per frame, Rate (events/ns), Cumulative events.
- 1 **Normalize**: per ns, per 100 frames, or none (disabled for Rate/Cumulative).
- 1 **Bin (ns)**: aggregates entries/exits into time bins (0 disables binning).
- 1 **Split axes** and **Exits negative** adjust visualization.
- 1 Actions: **Save Plot**, **Export Entry/Exit CSV**.

Axes:

- 1 X-axis: time in ns.
- 1 Y-axis: events per frame, events/ns, or cumulative events depending on mode.

Note: If **entries/exits** are missing, the GUI derives them from changes in **n_solvent** and warns in the status bar.

Plots tab

Contains three plots (tabs): Histogram, Matrix / Heatmap, Event Raster.

Histogram

Controls: **Histogram metric**, **Bins**, **Normalize**, **Log Y**, **Split zeros**, **Plot Histogram**, **Save Plot**, **Copy**.

Axes:

- 1 X-axis: selected metric (count or time in ns if **time**).
- 1 Y-axis: frame count or fraction of frames if normalized.

Notes:

- 1 Red line = mean; green line = median.
- 1 **Split zeros** separates zero vs non-zero frames and computes mean/median on the non-zero subset.

When this plot is misleading:

- 1 If most frames are zero occupancy, the histogram hides when occupancy occurs. Use Timeline or Event Raster.
- 1 Large stride compresses transient events and can bias the distribution.

Matrix / Heatmap

Controls: **Matrix source**, **Top solvents**, **Min occ %**.

Modes:

- 1 **Solvent occupancy (top N)**: binary presence matrix for the top N solvents by occupancy.
- 1 X-axis: time in ns.
- 1 Y-axis: solvent rank (row 0 = most occupied).
- 1 Color: 1 = present, 0 = absent.
- 1 **Residue hydration**: heatmap of **hydration_freq** per residue (requires residue hydration config).

Note: Heatmap requires per-frame solvent IDs (**analysis.store_ids=true**).

Event Raster

Controls: **Event stride**, **Segments**, **Min duration**.

Axes:

- 1 X-axis: time in ns.
- 1 Y-axis: solvent rank (same ordering as the heatmap).

Interpretation:

- 1 Points show occupancy events; **Segments** shows continuous occupancy spans.

When this plot is misleading:

- 1 High stride can hide short events or merge nearby events.
- 1 Small Top N can hide rare but important solvents.

Tables tab

- 1 **Filter**: text filter for row filtering.
- 1 Tabs: **Per Frame**, **Per Solvent**.
- 1 Tables are sortable; filtering applies to the active table.

Export step (data, report, extraction)

Export Data

- 1 Writes **metadata.json**, **soz_<name>/**, **bridge_<name>/**, and **hydration_<name>/** outputs to disk.
- 1 Uses the Output Settings directory unless overridden in the export dialog.

Export Report

- 1 Generates **report/report.md** and plots; **report/report.html** if **report_format=html**.
- 1 The Report panel shows the output path and copies it to the clipboard after export.

Extract tab (GUI)

Modes:

- 1 **Threshold**: rule-based filtering with operator and threshold.
- 1 **Percentile**: selects frames at/above a percentile of the metric distribution.
- 1 **Top N frames**: selects the top N frames by metric (ties may add frames).

Controls:

- 1 **SOZ, Metric, Operator, Threshold, Percentile, Top N frames**.
- 1 **Min run length** and **Gap tolerance** (disabled for Top N mode).
- 1 **Format**: currently **xtc**.
- 1 **Output directory** and **Use Output Settings directory** toggle.
- 1 **Preview** computes the rule and preview table; **Extract** writes outputs.

Outputs:

- 1 **<prefix>.xtc**: extracted trajectory.
- 1 **<prefix>_ref.pdb**: reference structure (first selected frame).
- 1 **<prefix>_frames.csv**: manifest (frame, time, n_solvent, optional solvent_ids hash).
- 1 **<prefix>_params.json**: extraction parameters.

Plot export and PowerPoint guidance

Export formats

Save Plot offers: PNG, SVG, EMF, PDF, CSV.

- 1 **PNG**: raster image.
- 1 **SVG**: vector (sanitized for PowerPoint edits).
- 1 **EMF**: Office-friendly vector, best with Inkscape installed.
- 1 **PDF**: vector, best with Inkscape installed.
- 1 **CSV**: plot data (timeline, histogram, heatmap, event raster, entry/exit).

PowerPoint tips

- 1 **SVG**: Insert > Pictures, then Ungroup (or Convert to Shape). If Ungroup fails, try a cut/paste cycle.
- 1 **EMF**: Insert > Pictures, then Group > Ungroup (accept conversion prompt).
- 1 For best fidelity, install Inkscape so SVG->EMF/PDF conversions use **--export-area-drawing**.

Plot CSV outputs (for reproducibility)

- 1 Timeline CSV: **time_ns**, metric column, optional secondary metric; includes **soz** when overlay is enabled.
- 1 Entry/Exit CSV: **time_ns**, **count**, **delta**, **entries**, **exits**, **entry_rate**, **exit_rate** (entries/exits follow current normalization/binning).
- 1 Histogram CSV: **bin_left**, **bin_right**, **bin_center**, **count** (fraction if normalized).
- 1 Heatmap CSV: **solvent_id** plus time columns (solvent occupancy) or hydration table columns for residue hydration.
- 1 Event Raster CSV: points (**solvent_id**, **row**, **frame**, **time_ns**) or segments (**solvent_id**, **row**, **start_frame**, **end_frame**, **start_ns**, **end_ns**, **duration_frames**).

Project JSON schema (current)

Minimal skeleton

```
{
  "version": "1.0",
  "inputs": {
    "topology": "path/to/topol.tpr",
    "trajectory": "path/to/traj.xtc"
  },
  "solvent": {
    "solvent_label": "Water",
    "water_resnames": ["SOL", "WAT", "TIP3", "HOH"],
    "water_oxygen_names": ["O", "OW", "OH2"],
    "water_hydrogen_names": ["H1", "H2", "HW1", "HW2"],
    "probe": {
      "selection": "name O OW OH2",
      "position": "atom"
    },
    "ion_resnames": ["NA", "CL"],
    "include_ions": false
  },
  "selections": {
    "selection_a": {
      "label": "selection_a",
      "selection": "protein and resid 10 and name CA",
      "require_unique": true
    }
  },
  "sozs": [],
  "analysis": {
    "frame_start": 0,
    "frame_stop": null,
    "stride": 1,
    "gap_tolerance": 1,
    "store_ids": true,
    "store_min_distances": true,
    "preview_frames": 200
  },
  "outputs": {
    "output_dir": "results",
    "write_per_frame": true,
    "write_parquet": false,
    "report_format": "md"
  },
  "extraction": {
    "rule": "n_solvent>=1",
    "min_run_length": 1,
    "gap_tolerance": 0,
    "output_format": "xtc",
    "output_dir": "results/extracted"
  },
  "bridges": [],
  "residue_hydration": []
}
```

Selections

- 1 Use **selections** (preferred) or legacy **seeds**.

- 1 Each selection supports either a full **selection** string or structured fields (**resid**, **resname**, **atomname**, **segid**, **chain**, **atom_indices**, **pdb_serials**).
- 1 **require_unique** enforces a single-atom match.
- 1 Bridges also accept legacy **seed_a** and **seed_b** keys.

SOZ logic tree

- 1 Node types: **distance**, **shell**, **and**, **or**, **not**.
- 1 Use **selection_label** in node params (legacy **seed_label** or **seed** are normalized).
- 1 Cutoffs accept **unit: "A"** or **unit: "nm"**.
- 1 Probe mode key is **probe_mode** (legacy **atom_mode** also accepted).

Allowed probe modes:

- 1 **probe**: use **solvent.probe.position**.
- 1 **atom**: use probe atoms directly.
- 1 **com**: center of mass of probe atoms per solvent residue (requires masses).
- 1 **cog**: center of geometry of probe atoms per residue.
- 1 **all**: all atoms in solvent residues.
- 1 Legacy **O** is treated as **atom**.

Solvent configuration

- 1 **water_resnames**: residue names that identify solvent molecules (UI shows "Solvent resnames").
- 1 Legacy alias: **solvent_resnames** is accepted in project JSON.
- 1 **probe.selection**: MDAnalysis selection for probe atoms within solvent residues.
- 1 **probe.position**: **atom**, **com**, or **cog** (global default).
- 1 **water_oxygen_names** and **water_hydrogen_names** are legacy fields kept for compatibility.
- 1 **solvent_label** is a display label only; set it to the actual solvent name for clarity (e.g., Methanol, DMSO).

Bridges

Each bridge uses two selections and two cutoffs.

- 1 **selection_a**, **selection_b**, **cutoff_a**, **cutoff_b**, **unit**, **probe_mode**.

Residue hydration

- 1 **residue_selection**: residues to score for hydration.
- 1 **cutoff**, **unit**, **probe_mode**.
- 1 **soz_name**: optional SOZ to use; default is the first SOZ.

Outputs and file map

Output map

Path	Meaning
-----	-----
metadata.json	Project snapshot, warnings, QC summary.
soz_<name>/per_frame.csv	Per-frame time series and events.
soz_<name>/per_solvent.csv	Per-solvent residence and occupancy summary.
soz_<name>/summary.json	Aggregated SOZ summary metrics.
soz_<name>/min_distance_traces.csv	Min distance traces per selection (if enabled).
bridge_<name>/per_frame.csv	Bridge-specific per-frame table.
bridge_<name>/per_solvent.csv	Bridge-specific per-solvent table.
hydration_<name>/hydration_table.csv	Residue hydration table.
report/report.md	Report markdown with figures.
report/report.html	Report HTML (if report_format=html).

per_frame.csv columns

- 1 **frame, time, n_solvent, entries, exits.**
- 1 **solvent_ids** (semicolon-delimited **resname:resid:segid**) if **analysis.store_ids=true**.
- 1 Optional min-distance columns (one per selection) if **analysis.store_min_distances=true**.

per_solvent.csv columns

- 1 **solvent_id, resindex, resname, resid, segid.**
- 1 **frames_present, occupancy_pct, entries.**
- 1 **mean_res_time_cont, median_res_time_cont.**
- 1 **mean_res_time_inter, median_res_time_inter.**

summary.json key fields

- 1 **n_frames, frame_stride, dt, time_unit.**
- 1 **occupancy_fraction:** fraction of sampled frames with **n_solvent > 0**.
- 1 **mean_n_solvent, median_n_solvent, max_n_solvent.**
- 1 **n_solvent_hist, entry_rate, exit_rate.**

Extraction (GUI + CLI)

Rule format

Rules are simple comparisons against per-frame metrics:

- 1 **n_solvent>=1**
- 1 **entries>=2**

1 **occupancy_fraction==1**

occupancy_fraction is computed per frame as 1 if **n_solvent>0**, else 0.

CLI extraction example

```
sozlab-cli extract \  
  --project /path/to/project.json \  
  --soz shell_only \  
  --rule "n_solvent>=2" \  
  --min-run 10 \  
  --gap 2 \  
  --out /path/to/extracted
```

Validation of extraction

1) Confirm **<prefix>_frames.csv** matches the expected number of frames. 2) Cross-check against **per_frame.csv** for the same SOZ. 3) Load **<prefix>_ref.pdb** and **<prefix>.xtc** in a viewer for spot checks.

Reproducibility and QC

- 1 Keep the project JSON and **metadata.json** for each run.
- 1 Record **analysis.frame_start**, **analysis.frame_stop**, **analysis.stride**, and **analysis.gap_tolerance**.
- 1 Use **sozlab-cli validate --project ... --max-frames 200** to compare fast vs slow evaluation.
- 1 Store preprocessing context in **inputs.processed_trajectory** and **inputs.preprocessing_notes**.

Troubleshooting (Symptom -> Confirm -> Fix)

Selection resolves to 0 atoms

Confirm: Selection Tester shows 0 hits. Fix: Verify resname/resid, add segid/chainID, or adjust selection scope.

Selection not unique but **require_unique** is checked

Confirm: Selection Tester shows multiple hits. Fix: Add segid/chainID/resid or disable **require_unique**.

Solvent resnames do not match any residues

Confirm: Project Doctor reports no solvent matches. Fix: Update **solvent.water_resnames** to match your topology.

Probe selection resolves to 0 atoms or misses residues

Confirm: Selection Tester -> Test Probe shows 0 hits or missing residues. Fix: Update **solvent.probe.selection** and **solvent.probe.position**.

Probe position com fails

Confirm: Error mentions missing masses. Fix: Add atom masses in the topology or use **probe.position: "cog"**.

Logic tree yields zero occupancy

Confirm: **per_frame.csv** has **n_solvent=0** for all frames. Fix: Check cutoffs and selection labels; use the Wizard Preview to test frame 0.

Heatmap/Event Raster says "No solvent IDs available"

Confirm: **per_frame.csv** lacks **solvent_ids**. Fix: Set **analysis.store_ids=true** (CLI: omit **--no-ids**).

Entry/Exit plot is flat

Confirm: **entries** and **exits** are zero or derived from **n_solvent**. Fix: Reduce stride, re-run with per-frame output enabled.

Per-frame tables are too large

Confirm: Disk usage is large or export is slow. Fix: Set **outputs.write_per_frame=false** (CLI: **--no-per-frame**).

Output directory not writable

Confirm: Logs show write errors. Fix: Update Output Settings or pass **--output** for CLI runs.

CLI command not found

Confirm: **sozlab-cli --help** fails. Fix: Activate the **sozlab** environment or run the installed script from your env.

SVG/EMF not editable in PowerPoint

Confirm: SVG opens as an icon or EMF is corrupted. Fix: Export again with Inkscape installed; use the PowerPoint tips above.

Appendix: keyboard shortcuts

- 1 **Ctrl+O**: Load project
- 1 **Ctrl+S**: Save project
- 1 **Ctrl+R**: Run analysis
- 1 **Ctrl+Shift+R**: Export report
- 1 **Ctrl+E**: Export data
- 1 **Ctrl+P**: Toggle presentation mode
- 1 **Ctrl+K**: Command palette

- 1 **Ctrl++** or **Ctrl+=**: Increase scale
- 1 **Ctrl+-**: Decrease scale
- 1 **Ctrl+0**: Reset scale
- 1 **Ctrl+Q**: Quit
- 1 **Ctrl+Mouse Wheel**: Adjust scale